DOCUMENT RESUME

ED 054 993 SO 001 744

AUTHOR Wulf, Christoph
TITLE Curriculum Evaluation.
PUB DATE 71
NOTE 29p.
JOURNAL CIT Zeitschrift Fur Padagogik; v17 n2 p175-201 1971

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Curriculum Evaluation; *Curriculum Research; *Evaluation Methods; *Models; Research Methodology; *Research Problems

ABSTRACT

In German, this article asserts that present concepts of evaluation are too narrow for curriculum evaluation. Research must differentiate different forms and roles of evaluation, produce a curriculum evaluation model, and resolve problems of values and judgement. The second section discusses the different forms and roles of evaluation: formative and summarative evaluation, micro- and macro evaluation, and intrinsic evaluation as opposed to sequential evaluation. The third section discusses models of curriculum evaluation as developed by 1) R.E. Stake, who defines three areas of evaluation--Hypotheses, Processes and Results; 2) M. Provus who defines a descrepancy model, and 3) D.L. Stufflebeam, who attempts to develop the structure of a design for evaluation. The fourth section discusses the problems of value and judgement in evaluation in relationship to 1; values, priorities, and norms; 2) the programatic and public-political sphere; and, 3) aspects of judgement and decision which are ambiguous in evaluation. The final section lists seven problems which need to be solved in order to develop theories and workable methods of curriculum evaluation. (Author)

ED054993

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

N. Helbourn

Zeitschrift
für
Pädagogik

Sonderdruck

Curriculum Evaluation
by Christoph Wulf.

SØ 001 744

Julius Beltz Weinheim

for Nick
with kind regards
Chris Wulf

Christoph Wulf
Deutsches Institut für
Internationale Pädagogische
Forschung
6 Frankfurt/Main 90
Schloßstraße 29

ED054993

CHRISTOPH WULF

Curriculumevaluation

I. Notwendigkeit einer Theorie der Curriculumevaluation[1]

In der Bundesrepublik findet heute Curriculumentwicklung an vielen Orten, zugleich aber auf verschiedenen Ebenen statt:

— Im Rahmen der *Lehrplanrevisionen und Richtlinienarbeit,* die von den Ländern betrieben wird[2],
— in vielen unmittelbar auf die Schule bezogenen *Initiativgruppen,* die an traditionellen Schulen und Gesamtschulen gebildet werden[3],
— als sorgfältige Entwicklung einiger *Teilcurricula,* wie sie an mehreren Universitäten und Forschungsinstituten geplant oder bereits erfolgt ist[4].

---

1 Für wertvolle Anregungen danke ich Dr. B. BLOOM, University of Chicago, Dr. R. STAKE, University of Illinois, Urbana, Professor Dr. G. PRIESEMANN, TU Berlin, und Dr. D. KAMPER, Universität Marburg.

2 Intensiv sind die Bemühungen um die Lehrplanrevision in Hessen, wo man das bestehende Schul- und Fächersystem durch die Entwicklung neuer Lehrpläne für die Klassen 5 bis 10 (Sekundarstufe I) überwinden will; sie sollen für die herkömmlichen Schulen und die Gesamtschulen gleichermaßen gültig sein. Die Kommission zur Reform der Hessischen Bildungspläne, Geschäftsstelle beim Erziehungswissenschaftlichen Seminar der Universität Marburg, gibt Berichte und Ergebnisse der Arbeit heraus. Vgl. den Bericht der Vorbereitenden Kommission unter Leitung von W. KLAFKI „Zur Lehrplanrevision für die Sekundarstufe in Hessen", den Bericht über die Curriculumtagung vom 18./19. 4. 1969 in Kassel/Reinhardswaldschule, die „Mitteilungen", Folge 1 (Nov. 1969), 2 (Febr. 1970), 3a und 3b (Mai 1970) und die Beispiele für lernzielorientierten Unterricht: Projekt 1: Urlaub, Projekt 2: Wohnen usw.

3 Übersicht über die unter dieser Rubrik zu nennenden Versuche zu erhalten, ist schwierig. Als Beispiel sei W. TESCHNER, Didaktik und Organisation des Deutschunterrichts an der Gesamtschule Berlin (BBR), Braunschweig 1968, genannt, ein Entwurf, der ein Curriculum für den Deutschunterricht des 7. Schuljahres enthält.

4 Vgl. D. KNAB, Ansätze zur Curriculumreform in der BRD, „betrifft erziehung", Febr. 1971, S. 15—28. Vgl.: Die Ausschreibung der Stiftung Volkswagenwerk für ein *C*urriculum *i*nstitutionalisierte *E*lementarerziehung (CIEL) oder die in Kürze erfolgende für ein *C*urriculum *ö*konomisch *s*ozio-*t*echnische Bildung (COST); die Curriculumarbeit des Instituts für die Pädagogik der Naturwissenschaften in Kiel; die Laborschule und das Oberstufenkolleg in Bielefeld mit ihren Plänen zur kontrollierten Curriculumentwicklung; das Schweizer Projekt EBAC, die „*E*ntwicklung und *b*egleitende *A*nalyse eines *C*urriculum"; die Arbeiten zur Curriculumrevision und Curriculumentwicklung am Institut für Bildungsforschung in der Max-Planck-Gesellschaft; das für die hier erörterte Problematik besonders interessante LOT-Projekt der Arbeitsgruppe für Unterrichtsforschung der Universität Konstanz, das von *l*ernziel*o*rientierten *T*ests ausgeht. Vgl. auch: H. BLANKERTZ, Ein Beispiel, Strategie zur Entwicklung des

Diese Bemühungen sind unterschiedlich zu bewerten, was ihren jeweiligen theoretischen Ansatz, den erreichten Grad der theoretischen Reflexion und die Qualität der verschiedenen Ergebnisse betrifft. Bereits in diesem für Deutschland noch frühen Stadium der Curriculumentwicklung sollte ein weiterer Schritt vollzogen werden: Es gilt eine Theorie der Curriculumevaluation zu entwerfen und für ihre Realisierung ein methodisches Instrumentarium zu entwickeln. Die systematische Erweiterung der Curriculumarbeit in dieser Hinsicht ist notwendig: ein neues Curriculum, über dessen Wert nichts zu erfahren ist, erscheint wertlos. Es bedarf aber nicht nur der kritischen Bewertung dessen, was als Produkt der Curriculumentwicklung entsteht, sondern auch der kritischen Beurteilung des ganzen Prozesses, in dem es entstand. So gesehen, ist Evaluation Begleitteil aller curricularen Arbeit überhaupt. Daher muß Evaluation von Anfang an in der ganzen Breite ihrer Aufgaben, ihrer Bedeutung, ihrer Verfahren begriffen werden, d. h. *die Theorie des Curriculum muß um die Theorie der Curriculumevaluation ergänzt werden.* Es besteht aber durchaus die Gefahr, daß der Begriff der Evaluation zu eng und oberflächlich verstanden wird, z. B. wenn Evaluation nur auf individuelle Leistungsmessung begrenzt wird[5], wodurch ihre wichtige Funktion für den Prozeß der Curriculumentwicklung übersehen wird. Dieser Gefahr sind tatsächlich große Teile der amerikanischen Curriculumentwicklung seit dem Anfang der sechziger Jahre erlegen. Sollen ähnliche Mißerfolge in der Bundesrepublik vermieden werden, dürfte zunächst ein Überblick über den augenblicklichen Stand der Evaluationsforschung in den USA angebracht sein; dadurch läßt sich ein guter Ausgangspunkt für die Entwicklung von Modellen, Methoden und einer Technologie der Curriculumevaluation gewinnen.

In den letzten Jahren ist Evaluation zu einem zentralen Thema der amerikanischen Curriculumforschung geworden[6]. Die Kontroversen über die Funktion

---

Lehrplans für das Fach „Arbeitslehre", in: Theorien und Modelle der Didaktik, München 1969, S. 172—185, und: F. ACHTENHAGEN/P. MENCK, Langfristige Curriculumentwicklung und mittelfristige Curriculumforschung, Z. f. P., 16. Jg. Nr. 3, 1970, S. 407—429.

5 Vgl. R. WOLFS scharfsinnige Satire über den oberflächlichen Gebrauch von Evaluationen: R. WOLF, A Model for Curriculum Evaluation, SRIS Quarterly (published by the School Research Information Service of Phi Delta Kappa International Headquarters), Vol. 3, No. 2, Bloomington 1970, S. 19—20.

6 Der entscheidende Grund für die Bedeutungszunahme der Evaluation liegt darin, daß die meisten die Curriculumreform finanzierenden Institutionen Evaluation als Bestandteil des Projektes fordern, seit der Elementary and Secondary Education Act von 1965 Evaluation als notwendigen Bestandteil der Entwicklung von Curriculumprojekten deklarierte.
Das wachsende Interesse an Evaluation verdeutlicht die gewaltig steigende Zahl der Publikationen zu diesem Thema. B. BLOOM setzte 1965 als Präsident der American Educational Research Association (AERA) eine Kommission für Curriculumevaluation ein. Ihr erstes Projekt war die AERA Monograph Series on Curriculum Evaluation; fünf Bände wurden zwischen 1967 und 1970 bei Rand McNALLY and Co., Chicago, publiziert. Curriculumevaluation erschien als Thema in drei von fünf Heften der 1969 Review of Educational Research (RER). Das Aprilheft 1970 der RER ist ganz dem Thema der „Educational Evaluation" ge-

der Evaluation haben wesentliche Beiträge zur Curriculumentwicklung geleistet[7]. Gekennzeichnet ist die Entwicklung der letzten Jahre durch eine zunehmende Vielfalt der Aufgaben, die der Evaluation zufallen. Dabei ist es zu grundlegenden Veränderungen in der Theorie und Methodologie gekommen, ohne daß es jedoch gelungen wäre, die dabei aufgeworfenen Probleme zu lösen. Summarisch läßt sich sagen: *Curriculumevaluation zielt auf die Sammlung, Verarbeitung und Interpretation von Daten mit dem Ziel, Entscheidungen über ein Curriculum zu fällen.* Das impliziert: „(1) objektive *Beschreibungen* von Zielen, Umwelt, Personal, Methoden und Inhalt und Ergebnissen; und (2) persönliche *Urteile* über die Qualität und Angemessenheit dieser Ziele, der Umwelt usw."[8].

SCRIVEN hat auf die Unterscheidung zwischen Ziel und Rolle der Evaluation aufmerksam gemacht[9]. Ziel ist die Sammlung von Informationen zum Zweck rational begründeter Entscheidungen über „etwas". Die Rolle der Evaluation hängt davon ab, was dieses „etwas" ist und von *wem* und *welche* Maßstäbe dabei angelegt werden. Aufgabe der Evaluation ist es z. B., zur Konstruktion eines Curriculum, zur Antizipation seines Erfolges oder zu seiner Verbesserung beizutragen. In der Regel erfolgt Evaluation für verschiedene Adressaten: z. B. Politiker, Fachleute, Lehrer, Eltern, Schüler und Curriculumentwickler. Mit Rücksicht auf ihr unterschiedliches Erkenntnisinteresse ist die *Form* und die *Sprache,* in der die Ergebnisse eines Evaluationsprozesses dargestellt werden, von Fall zu Fall zu variieren.

Eine Theorie der Evaluation muß:

— ein Vorstellungsschema liefern, das die Gebiete und Probleme klassifiziert, die evaluiert werden sollen,
— Strategien und Methoden der Evaluation entwickeln und
— ein System von Generalisationen für den Gebrauch der verschiedenen Verfahren und Techniken entwerfen[10].

---

widmet. Vgl. auch The Sixty-eighth Yeakbook of the National Society for the Study of Education Part II, Educational Evaluation: Nw Roles, New Means (edited by R. W. TYLER), The University of Chicago Press, Chicago 1969. In den hier genannten Veröffentlichungen finden sich umfassende Bibliographien zur Curriculumevaluation, die den Stand der Evaluationsforschung wiedergeben.

7 Vgl. RER, Vol. 39, No 3, 1969.

8 R. E. STAKE, Toward a Technology for the Evaluation of Educational Programs, Perspectives of Curriculum Evaluation, AERA Monograph Series on Curriculum Evaluation, No 1, Rand McNALLY and Co., Chicago 1967, S. 5.

9 M. SCRIVEN, The Methodology of Evaluation, Perspectives of Curriculum Evaluation, AERA Monograph Series on Curriculum Evaluation, No. 1, Rand McNALLY and Co., Chicago 1967, S. 39—83.

10 M. C. ALKIN, Evaluation Theory Development, UCLA Evaluation Comment, Center for the Study of Evaluation, Vol. 2, No. 1, Los Angeles, October 1969, S. 2. Ein Vorstellungsschema für den Bereich der Evaluation von Lernzielen könnte unter Heranziehung der Kategorien der Analyse der Lernzielproblematik erstellt werden, die DE CORTÉ für die vier Problemkreise (Formulierungsprobleme, Inventarisierungsprobleme, Klassifikations-

Eine solche Theorie, die eine Voraussage über die Angemessenheit verschiedener Evaluationsmethoden erlaubt und bereits die Voraussage bestimmter Ergebnisse ermöglicht, existiert noch nicht, wenngleich verschiedene Modelle entwickelt worden sind, die als ein erster Schritt zu einer Theorie der Evaluation anzusehen sind. So entwarf SCRIVEN bereits 1966 eine Methodologie der Evaluation[11] und STAKE 1967 ein Evaluationsmodell[12]. 1969 folgten ALKINS Evaluationstheorie[13], MAGUIRES Evaluationsmethodologie[14], PROVUS' Diskrepanzmodell[15] und STUFFLEBEAMS CIPP-Evaluationsmodell[16].

In mehreren dieser Evaluationsmodelle wird im Rahmen der Theorienbildung als Hauptaufgabe der Evaluation die *Bereitstellung von Daten für Entscheidungen* bezeichnet. Sie finden in Situationen statt, in denen zwischen mehreren Alternativen aufgrund unterschiedlicher Bewertung gewählt wird. Im Prozeß der Entwicklung und Einführung eines Curriculum in die Schule lassen sich mehrere Phasen kennzeichnen, in denen Entscheidungen von Entscheidungsträgern auf verschiedenen Ebenen zu fällen sind. Das bedeutet: es werden je

---

probleme, Wertprobleme) entworfen hat, obwohl sein Kategorienschema m. E. im einzelnen nicht immer überzeugend ist. E. DE CORTÉ, Analyse der Lernzielproblematik, Z. f. P., 17. Jg., Nr. 1, 1971, S. 75—89.

Aufgabe einer Theorie der Evaluation ist auch die Beantwortung der umstrittenen Frage, ob *Evaluation* eine eigene Disziplin ist, die sich von *Forschung* unterscheidet. Gegen eine prinzipielle Unterscheidung unter Betonung ihrer wechselseitigen Beziehungen ist I. WESTBURY, Curriculum Evaluation, RER, Vol. 40, No. 2, 1970, S. 252; ebenso dagegen: R. L. BAKER, Curriculum Evaluation, RER, Vol. 39, No. 3, 1969, S. 339—358; dafür: R. E. STAKE, Language Rationality and Assessment, Improving Educational Assessment and an Inventory of Measures of Affective Behavior (ed. by W. H. BEATTY), Association for Supervision and Curriculum Development (ASCD), National Education Association (NEA), Washington D. C. 1969, S. 14—40; STAKE sieht den Hauptunterschied zwischen Unterrichtsforschung, formativer Evaluation, summativer Evaluation und institutioneller Evaluation darin, daß der Grad an Verallgemeinerungsfähigkeit (Generalizability) sich verringert. J. K. HEMPHILL, The Relationships between Research and Evaluation Studies, The Sixty-eighth Yearbook of the NSSE, S. 189—220, bemüht sich, ebenfalls die Eigenständigkeit der Evaluation herauszuarbeiten. Entschieden für eine Unterscheidung zwischen Forschung und Evaluation treten A. G. LARKINS/J. P. SHAVER, Hard-Nosed Research and the Evaluation of Curricula, AERA Meeting, Los Angeles, February 7, 1969, ein. Möglicherweise bietet die für die Evaluation zentrale Aufgabe der Wertung und Entscheidungsvorbereitung einen fruchtbaren Ausgangspunkt für die Unterscheidung von Evaluation und Forschung in theoretischer und methodischer Hinsicht.

11 M. SCRIVEN, The Methodology of Evaluation, a. a. O.

12 R. E. STAKE, The Countenance of Educational Evaluation, Teachers College Record, Vol. 68, No. 7, 1967, S. 523—540.

13 M. C. ALKIN, Evaluation Theory Development, a. a. O.

14 T. O. MAGUIRE, Decisions and Curriculum Objectives, A Methodology for Evaluation, Alberta Journal of Educational Research, Vol. 16, 1969, S. 17—30.

15 M. PROVUS, Evaluation of Ongoing Programs in the Public School System, The Sixty-eighth Yearbook of the NSSE, a. a. O.

16 CIPP = *C*ontext *I*nput *P*rocess, *P*roduct; D. L. STUFFLEBEAM, Evaluation as Enlightment for Decision Making, Improving, Educational Assessment and an Inventory of Meassures of Affective Behavior (ed. W. H. BEATTY), Association for Supervision and Curriculum Development (ASCD), National Education Association (NEA), Washington D. C. 1969, S. 41—73.

nach Phase mehrere Funktionen der Entscheidung angesprochen und dementsprechend unterschiedliche Informationen benötigt; es müssen daher mehrere Evaluationsfelder unterschieden werden[17].

1. Erhebung der *Werte* und *Bedürfnisse* der Schüler und der Gesellschaft[18], um die Bildungs- und Lernziele auszuwählen, die in einem Curriculum in einer bestimmten historisch-gesellschaftlichen Situation realisiert werden sollen.
2. *Programmplanung* mit dem Ziel der Auswahl von Curricula. Hier gilt es, denen, die Entscheidungsbefugnis haben, möglichst objektive Informationen über verschiedene Curricula zu liefern, damit sie entscheiden können, welche speziellen Curricula ihren besonderen Bedürfnissen entsprechen.
3. *Programmimplementation:* Die Einführung curricularer Programme in die Schule mit der Aufgabe, festzustellen, in welchem Ausmaß die Intentionen der eingeführten Curricula verwirklicht werden und inwieweit sie den Bedürfnissen entsprechen, die als Voraussetzungen für die Entscheidung über die Einführung formuliert waren.
4. Evaluation der betreffenden Curricula mit dem Ziel, *sie zu verbessern, während sie noch im Prozeß der Entwicklung sind.*
5. Evaluation der Ergebnisse der betreffenden Curricula mit dem Ziel der *Programmbestätigung* (Programm Certification). Hier gilt es, eine abschließende Evaluation der Curricula vorzunehmen, um den generellen Wert und ihre Verwendbarkeit bestimmen zu können.

Auf die Evaluation im 5. Evaluationsfeld folgt wieder die Evaluation im ersten Feld, die auf z. T. veränderte Bedürfnisse und Interessen stoßen dürfte, wodurch ein neuer Ausgangspunkt für die Curriculumrevision gebildet wird.

Im ganzen bedeutet das, daß in jeder Phase der curricularen Arbeit evaluiert werden muß. Daraus folgt wieder, daß Rückkoppelung der Ergebnisse Einfluß haben muß auf die Gestaltung der jeweils nächsten Phase oder, um dieses gedankliche Modell zu benutzen, daß Evaluation in einem Regelkreis mit allen anderen Verfahren im Curriculumgefüge zusammengekoppelt ist. Ihre Funktion ist es, die Optimierung des Curriculum zu steuern.

Am Beispiel eines Teils der Evaluation, der Datenerhebung, exemplifiziert, heißt das:

---

17 M. C. Alkin, Evaluation Theory Development, a. a. O., S. 2 ff.

18 Vgl. dazu die zahlreichen Publikationen des National Assessment of Educational Progress; eine Übersicht über das National Assessment Program bietet J. C. Merwin und F. B. Womer, Evaluation in Assessing the Progress of Education to Provide Bases of Public Understanding and Public Policy, The Sixty-eighth Yearbook, a. a. O., S. 305—334. Vgl. dazu auch die Untersuchungen der International Association for the Evaluation of Educational Achievement (IEA); s. T. Husén, International Impact of Evaluation, The Sixty-eighth Yearbook of NSSE, a. a. O., S. 335—350, und W. Edelstein, Das „Projekt Schulleistung" im Institut für Bildungsforschung in der Max-Planck-Gesellschaft, Z. f. P., 16. Jg. Nr. 4, 1970, S. 517—529.

1. Daten über Bedürfnisse haben Einfluß auf die Bestimmung der Ziele.
2. Daten über die Gebiete, auf denen die Ziele liegen, haben Einfluß auf die Planung des Programms.
3. Daten über die schulische Einführung des Programms haben Einfluß auf die Einschätzung der Bedürfnisse, Ziele, Gebiete, Programmplanungen.
4. Die veränderte Einschätzung hat Einfluß auf die Umgestaltung des bisherigen Programms.
5. Daten über die Gesamtergebnisse des Programms haben Einfluß auf die vorläufig abschließende Einschätzung aller vorausliegenden Phasen und auf das Gesamturteil über das Programm und so fort.

Um die angestrebten Informationen optimal zu erhalten, bedarf es auch der Erarbeitung einer *Technologie der Evaluation*[19], die zu einer Theorie und Methodologie der Evaluation hinzutritt. Diese müßte Elemente der psychometrischen Testtechnologie[20], der sozialwissenschaftlichen Erhebungstechnologie, der Kommunikationstechnologie und der Unterrichtstechnologie einschließen. Das Fehlen einer solchen Technologie macht sich um so stärker bemerkbar, als die vorhandenen Tests sich selten für die Zwecke der Evaluation eignen. Die meisten von ihnen sind standardisierte Leistungstests, die zwar zur Feststellung von individuellen Leistungsunterschieden bei Schülern geeignet sind, nicht aber zur Evaluation eines Curriculum.

---

19 R. E. Stake, Toward a Technology for the Evaluation of Educational Programs, a. a. O.

20 Als Übersicht: vgl. O. Buros, The Sixth Mental Measurement Yearbook, The Gryphon Press, Highland Park, N. J. 1965.
Bloom hat versucht, einige Unzulänglichkeiten des Testens durch die Entwicklung einer *synthetischen Theorie*[1] des Testens zu beheben. Sie integriert *Messen* (Measurement) mit dem Ziel der Klassifikation, der Voraussage und des Experimentierens[2], *Evaluation* mit dem Ziel, Veränderungen in den Individuen festzustellen[3], *Schätzung* (Assessment) mit dem Ziel, die Charakteristika von Individuen in bezug zu einer speziellen Umgebung, Aufgabe oder Kriteriensituation zu erfassen[4]. Dabei wird Testen begriffen als „der Akt des Sammelns und Bearbeitens von Evidenz über menschliches Verhalten unter gegebenen Bedingungen zum Zeck des Verstehens, Voraussagens und Kontrollierens von zukünftigem menschlichen Verhalten"[5]. Die Grenze dieser Theorie liegt darin, daß sie nur eine Theorie des Testens ist. Es fehlt die Reflexion auf die verschiedenen Rollen der Evaluation, auf ihre Funktion für Entscheidungsprozesse, auf die für sie konstitutive Aufgabe, Wertungen und Urteile abzugeben. Evaluation ist umfassender als Testtheorie[6].

1 B. Bloom, Toward a Theory of Testing Which Includes Measurement-Evaluation-Assessment, Occasional Report No. 9, Center for the Study of Evaluation. University of California, Los Angeles, October 1968.

2 Ebenda, S. 4.

3 Ebenda, S. 5.

4 Ebenda, S. 9—12.

5 Ebenda, S. 23.

6 G. Glass, Comments on Professor Bloom's Paper Entitled „Toward a Theory of Testing Which Includes Measurement-Evaluation-Assessment". Occasional Report No. 11, Center for the Study of Evaluation, University of California, Los Angeles 1968.

Will man sich einen Überblick über das gesamte Problemfeld der Evaluation verschaffen, so lassen sich drei Problemkreise der Evaluationsforschung, die im folgenden zu behandeln sind, abgrenzen:
1. die verschiedenen Formen und Rollen der Evaluation,
2. die Versuche, ein Evaluationsmodell zu erstellen, das einen Rahmen für weitere Bemühungen um eine Theorie und Methodologie abgibt,
3. die Wert- und Entscheidungsprobleme der Evaluation und Methoden zu ihrer Lösung.

II. Rollen und Formen der Evaluation

Die erste und einfachste Unterscheidung zur Begriffs- und Funktionsbestimmung der Evaluation ist bereits getroffen. Es ist SCRIVENS Distinktion zwischen Zielen und Rollen. Dabei liegt für unsere Darstellung das Hauptgewicht zunächst auf den möglichen verschiedenen Rollen; denn als Ziel kann einheitlich die Absicht gesehen werden, Antworten auf bestimmte Fragen zu erhalten, die bestimmte Entscheidungen ermöglichen. Befragt werden muß alles, was in dem Gesamtprozeß der Curriculumentwicklung relevant ist, z. B.:
— die Personen, die mitarbeiten,
— die Ziele, die für ihre Arbeit gesetzt werden oder gelten,
— die Verfahren, mit deren Hilfe man die Ziele anstrebt,
— die Programme, von denen man sich dabei leiten läßt,
— die Instrumente, die dabei benutzt werden,
— die Urteile und Wertschätzungen, die die Arbeit steuern und die sich naturgemäß auf Personen, Ziele, Themen, Verfahren usw. beziehen können.

In Frage gestellt werden kann schließlich (auf einer weiteren Reflexionsstufe) die Evaluation selbst (und was aus ihr als Verfahren folgt oder gefolgt ist). Einige Klärungen innerhalb dieses weiten Problem-, Sach- und Begriffsfeldes sind bereits erfolgt. Von Belang sind m. E. die Distinktionen zwischen:
— formativer und summativer Evaluation,
— Mikro- und Makroevaluation,
— innerer und äußerer Evaluation,
— vergleichender und nicht vergleichender Evaluation,
— intrinsischer und Ergebnisevaluation.
Sie werden im folgenden kurz dargestellt[21].

21 Vgl. dazu auch: K.-H. FLECHSIG, Leitfaden zum Kolleg „Theorie des Unterrichts", Teil IV: Evaluation, Fachbereich Erziehungswissenschaft der Universität Konstanz, Konstanz 1970.

1. *Formative und summative Evaluation.* SCRIVENS Unterscheidung zwischen formativer und summativer Evaluation zur Kennzeichnung zweier Rollen von Evaluation hat Verbreitung gefunden. Formative Evaluation erfolgt, wenn das Curriculum sich noch im Prozeß der Entstehung (Formung) befindet. Summative Evaluation wird nach der Beendigung der Entwicklung eines Curriculum benötigt, wenn eine (vorläufig) abschließende Beschreibung und Wertung gewünscht wird. Entsprechend diesen beiden Rollen der Evaluation sind das Erkenntnisinteresse und die Methoden der Evaluation verschieden. CRONBACH hatte 1963 in seinem bekannten Aufsatz „Evaluation for Course Improvement"[22] die große Bedeutung der formativen Evaluation für die Curriculumentwicklung deutlich gemacht, ohne jedoch terminologisch zwischen formativer und summativer Evaluation zu unterscheiden. Die Aufgabe formativer Evaluation ist es, Schwächen der in der Entwicklung befindlichen Curricula aufzudecken und zu beseitigen[23]. Sie ist also als *der* Teil der Curriculumentwicklung zu begreifen, der Auskünfte über die Qualität des in der Entwicklung begriffenen Curriculum bietet. Aufgrund formativer Evaluation kann das Curriculum oder können einige seiner Teile revidiert werden, bis es seine Endfassung findet, die dann summativer Evaluation unterliegt. GROBMAN hat zu Recht darauf aufmerksam gemacht, daß „es keine scharf abgrenzbare Unterscheidung zwischen diesen beiden Phasen gibt; die formative Evaluation hört nicht auf, bevor die summative Evaluation anfangen kann"[24]. Es ist die Aufgabe formativer Evaluation, auf jede nur erdenkliche Weise Informationen über die Wirkungen des Curriculum zu erhalten; dabei sollten nicht nur die strengen Maßstäbe eines „Research Design" angelegt werden, vielmehr gilt es, einen Evaluationsplan zu entwerfen (s. Anm. 10). Um möglichst vielfältige Informationen über das Curriculum zu erhalten, schlägt CRONBACH statt der Verwendung von Tests, die auf individuelle Unterschiede zielen, die Einrichtung einer Gesamtheit von Items vor[25]. Aus einer beispielsweise 600 Items umfassenden Gesamtheit sollten besser je 50 Items 12 Gruppen von 50 Schülern gegeben werden als nur 50 Items 600 Schülern.

Die Entwicklung einer Gesamtheit von Items, die es erlauben, qualifizierte Aussagen innerhalb der formativen Evaluation zu erhalten, sollte mit der Formulierung von Lernzielen Hand in Hand gehen. Die Operationalisierung der Lernziele in Items legt ja obendrein eine permanente Revision der Lernziele und Items nahe. Es gilt die Übereinstimmung zwischen den formulierten Lernzielen, den impliziten und denen, die durch Items getestet werden, zu errei-

22 L. J. CRONBACH, Evaluation for Course Improvement, in: N. E. GRONLUND, Readings in Measurement and Evaluation, The Macmillan Co., COLLIER-MACMILLAN Ltd., London, 2nd Edition 1970, S. 37—52. 1. Fassung: Course Improvement Through Evaluation, Teachers College Record, Vol. 64, 1963, S. 672—683.

23 M. SCRIVEN, The Methodology of Evaluation, a. a. O., S. 40—43.

24 H. GROBMAN, Evaluation Activities of Curriculum Projects, AERA Monograph Series on Curriculum Evaluation, No. 2, Rand McNALLY and Co., Chicago 1968, S. 14.

25 L. J. CRONBACH, Evaluation for Course Improvement, a. a. O., S. 46.

chen. Dabei ist das Validitätsproblem schwierig zu lösen. Es ist notwendig, Lernziele und Curriculuminhalte, Lernziele und Inhalte der Prüfung, Curriculuminhalt und Prüfungsinhalt so zusammenzubringen, daß sie wechselseitige Identifikation zulassen[26]. Die Herstellung einer derartigen Entsprechung ist eine wichtige Aufgabe des Evaluierenden, der er sich unterziehen muß, bevor er das Curriculum im Rahmen der formativen Evaluation hinsichtlich seiner Wirkungen auf Schüler mit dem Ziel der Verbesserung evaluieren kann.

LINDVALL und Cox formulieren vier Fragen, die die Funktion von Kategorien für formative Evaluation haben; jede Kategorie enthält mehrere Unterkategorien, die ebenfalls in Form von Fragen entworfen werden. Mit der Verwendung dieses *Kategoriensystems* kann formative Evaluation beginnen. Es wurde von den Autoren vor allem im Hinblick auf die formative Evaluation der *Individually Prescribed Instruction* entwickelt, beansprucht aber zugleich allgemeine Verwendungsfähigkeit:

1. Welche Ziele sollte das Programm erreichen?
   a) Sind die formulierten Ziele wirklich Lernziele?
   b) Sind die formulierten Lernziele die wirklichen Ziele des Programms?
   c) Sind die Ziele wertvoll?
   d) Sind die Ziele erreichbar?
2. Welches ist der Plan, diese Lernziele zu erreichen?
   a) Verspricht der Plan, zur Erreichung der Lernziele beizutragen?
   b) Ist der Plan genügend detailliert entwickelt?
   c) Werden Plan und Verfahren mühelos von den Leuten verstanden werden, die sie verwirklichen sollen?
   d) Ist es wahrscheinlich, daß der Plan verwirklicht werden kann?
3. Stellt das funktionierende Programm eine richtige Einführung des Plans dar?
   a) Welches sind die spezifischen Punkte, die in einer Analyse der Funktion zu beobachten sind?
   b) Werden die Aktivitäten wirklich entsprechend dem Plan ausgetragen?
   c) Wie kann erreicht werden, daß Plan und Verwirklichung korrespondieren?
   d) Führt die Untersuchung der wirklichen Funktion zu irgendwelchen Veränderungsvorschlägen?
4. Erreicht das Programm, wenn es entwickelt und verwirklicht ist, sein erwünschtes Ziel?
   a) Sehen die Pläne die Evaluation aller Programmziele vor?
   b) Sind die Evaluationsverfahren reliabel?
   c) Ist der gesamte Evaluationsprozeß umfassend genug, um das benötigte Gesamtbild der Programmergebnisse zu liefern?
   d) Welches sind die Folgerungen aus den Ergebnissen für die Modifikation des Programms[27]?

26 M. SCRIVEN, The Methodology of Evaluation, a. a. O., S. 59.

Summative Evaluation, auch Ergebnisevaluation oder Produktevaluation genannt, ist die abschließende Evaluation eines Curriculum[28]. SCRIVEN hat ihre prinzipielle Gleichwertigkeit mit formativer Evaluation betont. Ihre Rolle ist es, nach der Evaluation des Curriculum Daten zur Verfügung zu stellen, die seinen Wert erkennen lassen und die „den Schulsystemen helfen, Entscheidungen über die Adaptation und den Gebrauch von Materialien zu treffen"[29]. Evaluation kann dabei einmal auf das Curriculummaterial selbst bezogen werden und seine Validität im Hinblick auf seine Gestaltung und auf seine Lernziele prüfen. Sie kann aber auch das Verhalten der Schüler untersuchen, die mit diesem Curriculummaterial arbeiten, um festzustellen, welche Fähigkeiten und welche Veränderungen im Verhalten sie erreicht haben.

2. Mikro- und Makroevaluation. Innerhalb einer abschließenden Evaluation sind zwei Schwerpunkte möglich. Den einen bildet die *Mikroevaluation,* also die Evaluation von Einzelteilen unter einer bestimmten Fragestellung, den anderen die *Makroevaluation,* die auf eine Beantwortung von allgemeinen Fragen, z. B. über die Benutzbarkeit des Materials, zielt. Summative Evaluation wird im allgemeinen auf die Kombination von Mikroevaluation und Makroevaluation zielen, um zu einer fundierten abschließenden Beschreibung und Wertung des Curriculum zu kommen.

3. Innere und äußere Evaluation. Formative Evaluation erfolgt als Teil der Entwicklung eines Curriculum. Sie wird am besten von den an der Entwicklung Beteiligten ausgeführt (innere Evaluation). Da sie in der Entwicklungsarbeit stehen, haben sie Kenntnisse vom Gegenstand der Evaluation. In einem stärkeren Maße mit dem Projekt identifiziert als Evaluierende, die nicht mit der Projektherstellung verbunden sind, werden sie nicht so leicht durch formalen Dogmatismus den kreativen Schwung der übrigen Curriculumhersteller bremsen. Auch können sie eher bei der rechtzeitigen deutlichen Formulierung der Lernziele helfen[30]. Der dynamische Charakter der Curriculumentwicklung for-

27 C. M. LINDVALL/R. C. COX, Evaluation as a Tool in Curriculum Development, The IPI Evaluation Program, AERA Monograph Series on Curriculum Evaluation, No. 5, Rand McNally and Co., Chicago 1970; die Kategorien sind den S. 5 bis 11 entnommen. Diese Monographie verdient trotz mehrerer kritischer Einwände Interesse, da sie den Versuch macht, ein System der Evaluation für den umfassenden innovativen Schulversuch mit Individually Prescribed Instruction zu entwerfen. Dabei liegen die Schwerpunkte neben formativer und summativer Evaluation auf der Evaluation des Schülerverhaltens in IPI. Über IPI vgl. die zahlreichen Veröffentlichungen des Learning Research and Development Center der Universität Pittsburgh und die des Research for Better Schools Laboratory in Philadelphia.

28 Vgl. auch M. SCRIVEN, The Methodology of Evaluation, a. a. O., S. 72—82, in der versucht wird, in Anlehnung an BLOOMS Taxonomie (B. BLOOM et al. Taxonomy of Educational Objectives, The Classification of Educational Goals, Handbook I: Cognitive Domain, David McKay Company, New York 1956) Leistungskriterien für Evaluationsstudien aufzustellen.

29 H. GROBMAN, Evaluation Activities of Curriculum Projects, a. a. O., S. 14.

30 C. M. LINDVALL/R. C. COX, The IPI Evaluation Program, a. a. O., S. 4.

dert eine schnelle Anpassungsbereitschaft an die sich wandelnden Fragen und Probleme der Curriculumhersteller, die nur ein mit dem Projekt mitarbeitender Evaluierender leisten kann. Bei formativer Evaluation ist der Zeitpunkt ihres Beginns und die Art ihrer Durchführung zu bedenken. ATKIN hat auf diesbezügliche Probleme hingewiesen[31]. So kann z. B. das Testen eines schwierigen Vorstellungszusammenhangs zu einem zu frühen Zeitpunkt negative Folgen im Hinblick auf ein weiteres vertieftes Verständnis dieses Zusammenhangs haben.

Summative Evaluation kann von Evaluierenden geleistet werden, die nicht in Verbindung mit Curriculumentwicklung oder -realisierung stehen (äußere Evaluation). Das hat den Vorteil, daß die Evaluationskriterien oft objektiver gewählt werden, mithin auch größere Objektivität bei umfassenden Wertungen über das Curriculum gegeben ist. Auch ist es für nicht an der Projektentwicklung beteiligte Evaluierende leichter, Vergleichsuntersuchungen verschiedener Curricula vorzunehmen. Ein solches Evaluationsteam kann sich aus verschiedenen Spezialisten zusammensetzen, so daß ein großes Spektrum von methodischen Ansätzen gewährleistet wird.

4. Vergleichende und nicht vergleichende Evaluation. Betrachtet man die Ergebnisse der Evaluation neuer Curricula in den USA, die im Vergleich zu Kontrollgruppen erfolgt ist, die mit altem Unterrichtsmaterial arbeiten, fällt im allgemeinen das gute Abschneiden der neuen Curricula und die Unzulänglichkeit der alten Materialien ins Auge; zuweilen allerdings ergeben sich keine signifikanten Unterschiede. Das hat u. a. dazu geführt, daß die Zulänglichkeit der Kriterien des Vergleichs angezweifelt wurde, da sie vor allem im Hinblick auf die Lernziele der neuen Curricula erstellt wurden und für vergleichende Evaluation nicht angemessen waren. Infolgedessen hat man oft dem Urteil von Fachleuten über Inhalt und Ziele und Unterrichtsprozesse mehr Wert zugemessen als empirisch festgestellten geringen Unterschieden von Schülern in Experimentier- und Kontrollgruppen[32]. CRONBACH schlägt die genaue Untersuchung des „Verhaltens einer genau beschriebenen Gruppe in bezug auf viele wichtige Lernziele und Nebeneffekte nach dem Kursus“[33] vor. Durch eine genaue Evaluation des Verhaltens *einer* Gruppe glaubt CRONBACH zuverlässigere Daten zu erhalten als durch den methodisch schwierigen Vergleich verschiedenartig unterrichteter Gruppen. SCRIVEN wendet sich gegen diese Annahme, indem er darauf hinweist, daß die Evaluation eines Curriculum nicht nur die Beschreibung impliziert, sondern auch seine Wertung im Hinblick auf andere einschließt[34]. Ob

31 J. M. ATKIN, Some Evaluation Problems in a Course Content Improvement Project, Journal of Research in Science Teaching, Vol. 1, 1963, S. 129—132.
32 Vgl. dazu: L. CRONBACH, Evaluation for Course Improvement, a. a. O., S. 44.
33 Ebenda, S. 43.
34 M. SCRIVEN, The Methodology of Evaluation, a. a. O., S. 64.

vergleichende oder nicht vergleichende Evaluation gewählt wird, richtet sich nach der Ausgangssituation und dem Erkenntnisinteresse. Evaluation mit Hilfe des Vergleichs empfiehlt sich dann, wenn die allgemeinen Zielsetzungen gleich sind, d. h. wenn z. B. zwischen zwei Curricula oder zwei Teilen eines Curriculum gewählt werden soll, die auf verschiedenen Wegen das gleiche Ziel erreichen wollen. Eine nicht vergleichende Evaluation ist dann angebracht, wenn die Unterschiede zwischen den Curricula aufgrund verschiedener Zielsetzungen so groß sind, daß die Summe der Gemeinsamkeiten zu klein ist.

5. *Intrinsische Evaluation und Ergebnisevaluation.* Unter intrinsischer Evaluation wird die Evaluation von beispielsweise Inhalten, Lernzielen, Zensierungsverfahren verstanden, ohne daß ihre Auswirkungen auf die Schüler untersucht werden. Ihre Kriterien sind im allgemeinen nicht operationalisiert und beziehen sich auf den zu evaluierenden Gegenstand selbst[35]. Die Schwierigkeiten bei der intrinsischen Evaluation liegen darin, daß sie die Kriterien und Ziele formulieren muß, durch die Inhalt, Lehrerverhalten usw. evaluiert werden, zumal die von den Curriculumentwicklern formulierten Lernziele und Kriterien oft nicht gleich den impliziten Lernzielen und Kriterien sind[36].

Unter der reinen Ergebnisevaluation wird dagegen die Untersuchung der Effekte des Curriculum auf die Schüler verstanden. Sie werden z. B. durch Vor- und Nachtest in der Experimentiergruppe festgestellt. Dabei wird oft davon ausgegangen, daß „es nicht nur ein irrelevantes, sondern auch ein keineswegs reliables Vorgehen ist, die Erhebung von Lernzielen oder Inhalten in die Evaluation eines Kursus hineinzubringen"[37]. Für diese Art von Evaluation ist es weniger wichtig, was die Lehrer und Schüler als ihr Tun *bezeichnen,* wichtig ist, was sie *wirklich* tun. Die Schwierigkeiten bei der reinen Ergebnisevaluation liegen darin, *„mit einem nützlichen Allgemeinheitsgrad* zu beschreiben, was der Schüler gelernt hat"[38].

Aufgabe der Evaluation ist es jedoch nicht nur, die Ergebnisse eines Curriculum festzustellen; sie muß auch die Frage beantworten, wie diese im Verhältnis zu anderen Curricula zu werten sind. Dazu bedarf es der Spezifizierung des Lernerfolges, seiner Begründung und der Formulierung von Zielen und Kriterien. In der Ergebnisevaluation muß entschieden werden, „welches Verhalten zu werten ist als adäquates Verständnis für Schüler auf einem bestimmten Niveau

---

35 M. Scriven, The Methodology of Evaluation, a. a. O., S. 53.

36 Vgl. als Beispiel für ein System intrinsischer Curriculumevaluation: I. Morrissett/W. W. Stevens, Curriculum Analysis, Social Education, Vol. 31, 1967, S. 483—487.
Siehe auch: I. Morrissett, W. W. Stevens and C. P. Woodley, A Model for Analyzing Curriculum Materials and Classroom Transactions, Social Studies Curriculum Development, 39th Yearbook (ed. by D. Fraser), National Council for the Social Studies, NEA, Washington D. C. 1969, S. 229—276.

37 M. Scriven, The Methodology of Evaluation, a. a. O., S. 59.

38 M. Scriven, The Methodology of Evaluation, a. a. O., S. 60.

für ein bestimmtes Fach, und es bedarf der Anwendung dieser Entscheidung auf Daten über das spätere Verhalten der Schüler, um eine Gesamtevaluation zu ermöglichen"[39]. Es muß also das Verhalten der Schüler analysiert werden, um zu bestimmen, wo Unzulänglichkeiten im Hinblick auf bestimmte Kriterien oder Lernziele liegen. Im allgemeinen verbindet man intrinsische und Ergebnisevaluation, um zu möglichst umfassenden Ergebnissen zu kommen.

III. Modelle der Curriculumevaluation

1. Das Stake*sche Evaluationsmodell.* Einen Schritt weiter in die Problematik der Evaluationstheorie führt der Entwurf von Modellen für die Curriculumevaluation. Sie liegen auf einer höheren begrifflichen Stufe als die Distinktionen zur Beschreibung unterschiedlicher Evaluationsformen und Evaluationsfunktionen. Geht man wieder davon aus, daß das Verfahren der Evaluation beabsichtigt, relevante Informationen für die Verbesserung von Curricula zu erhalten, so sind vor allem drei Felder wichtig:

— das Vorfeld des Unterrichts,
— das Aktionsfeld des Unterrichts,
— das Feld seiner Wirkungen.

Etwas detaillierter beschrieben, heißt das: Das erste Evaluationsfeld enthält alle Bedingungen, die vor dem Unterricht bestehen und Einfluß auf seine Ergebnisse haben. Zu diesen gehört die Erhebung der sozialen Situation der Schüler, ihrer Bedürfnisse und Interessen und ihrer Umwelt mit ihren Wert- und Zielvorstellungen[40].

Zum zweiten Evaluationsfeld gehören alle Bedingungen innerhalb der Schule, die für die Erreichung der Ergebnisse wichtig sind. Es schließt den Prozeß der Einbringung des Curriculum in die Schule ein, die Einführung des Lehrers in die Arbeit mit den curricularen Materialien und vor allem die Unterrichtspro-

---

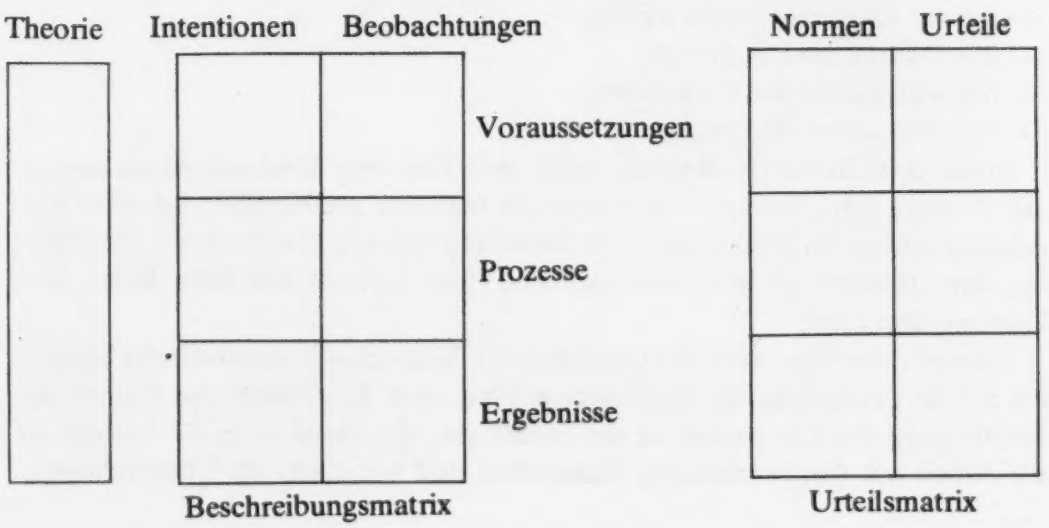
39 M. Scriven, The Methodology of Evaluation, a. a. O., S. 61.

40 Will man Instrumente zur Erhebung von Daten in diesem Bereich erstellen, bieten folgende amerikanische Werke wertvolle Hilfe:
E. E. Maccoby/N. Maccoby, The Interview: A Tool of Social Science, Handbook of Social Psychology (ed. by G. Lindzey), Vol. 1, Chapter 12, Addison-Wesley, Cambridge 1954, S. 449—487.
Ch. F. Cannell/R. L. Kahn, Interviewing, Handbook of Social Psychology (ed. by G. Lindzey and E. Aronson), Vol. 2, Chapter 15, Addison-Wesley, Reading, Mass. 1968, S. 526—595.
H. Hyman, Survey Design and Analysis: Principles, Cases, and Procedures, The Free Press, Glencoe, Ill. 1955.
P. F. Lazarsfeld/M. Rosenberg, The Language of Social Research, The Free Press, Glencoe, Ill. 1955.
C. Selltiz et al., Research Methods in Social Relations, Henry Holt, New York 1959.
A. N. Oppenheim, Questionnaire Design and Attitude Measurement, Basic Books, New York 1966.

zesse[41]: die Interaktionen der Lehrer mit den Schülern, der Schüler untereinander und der Schüler mit dem Unterrichtsmaterial. Das dritte beinhaltet die augenblicklichen, mittelfristigen und langfristigen Ergebnisse des gesamten Unterrichts[42].

STAKE hat 1967 die drei Evaluationsfelder *Voraussetzungen, Prozesse, Ergebnisse* genannt und eine verstärkte Evaluation der Voraussetzungen und der Prozesse gefordert[43]. Zu diesem Zweck entwickelt er eine Evaluationsmatrix, indem er über die drei Evaluationsfelder zwei Beschreibungskategorien, „Intentionen" und „Beobachtungen", und zwei Urteilskategorien, „Normen" und „Urteile", legt. Dadurch entsteht eine zwölffeldige Evaluationsmatrix. Voraussetzung für ihre Anwendung ist die Kenntnis der Theorie des Curriculum, dessen Intentionen mit ihrer Hilfe zu erfassen sind.

Evaluationsmatrix nach STAKE[44]



41 Vgl. die bibliographischen Angaben in: D. D. SJOGREN, Measurement Techniques in Evaluation, RER, Vol. 40, No. 2, 1970, S. 303—307; vgl. Anm. 69.

42 Vgl. die bibliographischen Angaben, ebenda, S. 307—312.

43 R. E. STAKE, The Countenance of Educational Evaluation, a. a. O., hier verwendet: ein Nachdruck des Artikels mit den Seiten 1—11; zur Übersetzung: Antecedents = Voraussetzungen; Transactions = Prozesse; Outcomes = Ergebnisse; vgl. dazu auch: ders., Language, Rationality, and Assessment, a. a. O. Ein ganz genaues Instrumentarium zur Erprobung und Bewertung von Unterrichtsprozessen gibt es z. Z. noch nicht. Insbesondere ist die Funktion der Unterrichtssprache in Unterrichtsprozessen weitgehend unerforscht; vgl. G. PRIESEMANN, Zur Theorie der Unterrichtssprache, Düsseldorf 1971.

44 R. E. STAKE, The Countenance of Educational Evaluation, a. a. O., S. 4.

Unter „Intentionen" werden z. B. die geplante Erarbeitung von Sachgebieten oder das erwartete Schülerverhalten verstanden. Sie sollten möglichst präzise verbalisiert werden. Die Diskrepanz zwischen Lernzielen und Lernergebnissen ist nicht ohne weiteres Zeugnis für einen negativen Verlauf des Unterrichts. Es können unvorhergesehene Ergebnisse erzielt worden sein, die durchaus positiv zu werten sind. Ob Lernziele *immer* vor dem Unterricht festgelegt und ob sie *immer* in Verhaltenstermini bestimmt werden müssen, ist in der amerikanischen Curriculumforschung umstritten[45]. „Beobachtungen" bilden die zweite Beschreibungskategorie der Matrix. Hier geht es um die Beobachtung dessen, was in den drei Evaluationsfeldern beschreibenswert erscheint. Es setzt die Auswahl von Kriterien voraus, unter denen zu beobachten ist. Diese ist eine Entscheidung des Evaluierenden, bei der es im Curriculum formulierte Kriterien und von außen herangetragene zu berücksichtigen gilt. Es bieten sich zwei Wege an, die Daten der beiden Beschreibungskategorien zu verarbeiten: *„Die Kontingenzen zwischen Voraussetzungen, Prozessen und Ergebnissen zu finden und die Kongruenz zwischen Intentionen und Beobachtungen zu finden"*[46]. Die Kontingenzen zwischen den einzelnen Variablen der drei Evaluationsfelder bedürfen der Beachtung: „Insofern, als Evaluation die Suche nach Beziehungen ist, die eine Verbesserung der Erziehung erlauben, ist es Aufgabe des Evaluierenden, Ergebnisse zu identifizieren, die mit bestimmten Voraussetzungen und Unterrichtsprozessen kontingent sind"[47].

Zwischen den drei Evaluationsfeldern besteht in der ersten Beschreibungskategorie, „Intentionen", eine logische Kontingenz, im Bereich der zweiten, „Beobachtungen", eine empirische Kontingenz. Vollständige Kongruenz der Daten eines Curriculum besteht, wenn alle beabsichtigten Voraussetzungen, Prozesse und Ergebnisse eintreten. Das ist selten der Fall. Im allgemeinen wird der

---

45 Zu dieser für die Curriculumentwicklung und Curriculumevaluation wichtigen Streitfrage vgl.: RER, Vol. 39, No. 3, 1969; besonders dazu: Instructional Objectives, AERA Monograph Series on Curriculum Evaluation, Vol. 3, Rand McNally and Co., Chicago 1969. Hierin der Streit zwischen E. W. Eisner auf der einen Seite, der außer für „Instructional Objectives" auch für „Expressive Objectives" eintritt, die die pädagogische Situation, Probleme und ähnliches angeben, aber ihre Ergebnisse nicht vorwegnehmen, und W. J. Popham and H. J. Sullivan auf der anderen Seite, die derartige Lernziele nicht als gleichwertig neben den operationalisierten „Instructional Objectives" anerkennen wollen. Wenn man „Expressive Objectives" als eine zulässige Art, Lernziele zu formulieren, anerkennt, ergibt sich als Aufgabe der Evaluation, „nicht eine gemeinsame Norm auf die hervorgebrachten Produkte anzuwenden, sondern die Aufgabe, das zu reflektieren, was produziert wurde, um seine Einzigartigkeit und Bedeutung zu enthüllen. Im expressiven Kontext ist das Produkt wahrscheinlich genauso überraschend für den Hersteller, als es für den Lehrer ist, dem es begegnet." (E. Eisner, ebenda, S. 16.)
R. E. Stake unterstützt diese Auffassung teilweise, Language Rationality and Assessment, a. a. O., S. 28 ff.
Evaluation wird heute durch diese Position vor neue, bisher kaum reflektierte Aufgaben gestellt.

46 R. E. Stake, The Countenance of Educational Evaluation, a. a. O., S. 6.

47 Ebenda

Evaluierende Kongruenz lediglich bei einigen Teilzielen feststellen können. Kongruenz besagt nur, daß das, was beabsichtigt war, eintrat, nicht jedoch, daß die Ergebnisse *reliabel* oder *valide* sind.

Die Urteilsmatrix besteht aus zwei Kategorien, den *„Normen"* und den *„Urteilen"*. Mit welchen Methoden und Strategien die Normen zu erstellen sind, ist ein zentrales Problem der Evaluation. Normen sind im allgemeinen gebunden an die historisch-gesellschaftliche Situation. Mit den Lernzielen können sie nicht gleichgesetzt werden, wie es in einigen Evaluationsprojekten geschehen ist, da dann Evaluation nur Überprüfung der Erfüllung von Lernzielen wäre. Vielmehr müssen auch die Lernziele an Normen gemessen werden. Zwei Arten von Normen sind denkbar, *absolute* und *relative*. Letztere sind beispielsweise in einem alternativ vorliegenden Curriculum enthalten. Wenigstens an einem der beiden Normengefüge muß das Curriculum durch Urteilen gemessen werden, um seinen Wert zu bestimmen.

Das Provus'sche *Diskrepanzmodell.* Provus macht darauf aufmerksam, daß die entscheidenden evaluativen Informationen durch die Feststellung von Diskrepanz gewonnen werden. Danach besteht der Prozeß der Evaluation darin,

— Normen festzulegen,
— die Diskrepanz zwischen dem Curriculum und den Normen festzustellen und
— die Diskrepanzinformation dazu zu benutzen, die Schwächen des Curriculum zu identifizieren[48].

Die Evaluation eines Curriculum soll vier Entwicklungsstadien durchlaufen und dabei drei größere Inhaltskategorien einschließen. Das erste Entwicklungsstadium ist die *Definition* des Curriculum, das zweite seine *Einführung* in die Schule (Installation), das dritte die curricularen *Prozesse,* das vierte seine Fertigstellung als *Produkt.* Nach diesen vier kann als fünftes Entwicklungsstadium die *Kosten-Ertrags-Berechnung*[49] erfolgen. In den vier bzw. fünf Entwicklungsstadien „bleiben die inhaltlichen Spezifikationen dieselben: *Eingabe, Prozeß* und *Ergebnis,* relativ zu Zeit und Geld"[50].

Unter *Definition* eines Curriculum wird seine genaue Beschreibung im Hinblick auf seine Lernziele, die Schüler, Lehrer, Medien und den Prozeß verstanden, der zur Erreichung der Lernziele führt. Sie dient als Norm zur Evaluation

48 M. Provus, Evaluation of Ongoing Programs in the Public School System, a. a. O., S. 245.
49 Im Deutschen werden nach Billerbeck die Begriffe Cost-Benefit Analysis und Cost-Effectiveness Analysis unter Kosten-Ertrags-Analyse subsumiert; (K. Billerbeck, Kosten-Ertrags-Analyse, Berlin 1968). Zur Unterscheidung des hier vorliegenden Begriffs Cost-Benefit Analysis von Cost-Effectiveness Analysis vgl. M. Blaug, Cost-Benefit and Cost-Effectiveness Analysis of Education, Meeting of Ad Hoc Group on Budgeting Program Analysis and Cost-Effectiveness in Educational Planning (Directorate for Scientific Affairs, Educational Management Techniques) OECD, Paris, 30th January 1968.
50 M. Provus, Evaluation of Ongoing Programs in the Public School System, a. a. O., S. 248.

des Curriculum. Werden im (zweiten) Entwicklungsstadium, der *Einführung* des Curriculum, Informationen gewonnen, die eine Diskrepanz zwischen den Erwartungen hinsichtlich der Einführung des als Norm definierten Curriculum und dem tatsächlichen Lehrer- und Schülerverhalten anzeigen, muß entweder das Verhalten so korrigiert werden, daß es der Norm entspricht, oder es muß die Norm verändert werden.

In dem (dritten) Entwicklungsstadium, dem *Prozeß,* bildet der Teil der Curriculumdefinition die Norm, der die Intentionen hinsichtlich des Unterrichtsprozesses enthält. Wird eine Diskrepanzinformation zwischen Intentionen und tatsächlichem Prozeß erhalten, muß sie zur Neudefinition der Intentionen oder zu einer besseren Durchführung des Unterrichtsprozesses führen. Im (vierten) Entwicklungsstadium, in dem es um das *Produkt* des Unterrichts geht, gilt es festzustellen, ob das Curriculum seine in der Definition als Norm gesetzten End- und Gesamtlernziele erreicht hat oder ob Diskrepanz festgestellt werden muß.

Mit diesem Evaluationsmodell wird die Evaluationskategorie der *Diskrepanz* gewonnen. Darüber hinaus bietet PROVUS brauchbare Hinweise auf die methodischen Probleme seiner Anwendung und die spezifischen Aufgaben von Evaluationsgruppen, die hier nicht mehr berücksichtigt werden können[51].

Da es ein Modell für die Evaluation laufender curricularer Programme ist, fehlt die Reflexion auf die Aufgaben der Evaluation zur „Erhebung der Werte und Bedürfnisse". Es fehlt auch die Reflexion auf die Notwendigkeit einer auf den Ergebnissen der Produktevaluation fußenden Revision des Curriculum. Außerdem werden die in den einzelnen Entwicklungsphasen notwendigen Entscheidungen nicht als Probleme der Wertung innerhalb der Evaluation gesehen.

3. Das STUFFLEBEAM*sche Evaluationsmodell.* Gerade die Transparenz von Entscheidungsprozessen bildet das Schwergewicht der Bemühungen STUFFLEBEAMS[52]. In diesem Rahmen werden sieben Entscheidungsvariablen unterschieden, die in den verschiedenen Entscheidungssituationen heranzuziehen sind:

— die Ebene der Entscheidungen,
— der Konzentrationspunkt der Entscheidungen,
— die Substanz der Entscheidungen,
— die Funktion der Entscheidungen,
— die Gegenstände der Entscheidungen,
— der Zeitpunkt der Entscheidungen,
— der Grad an kritischer Reflektiertheit der Entscheidungen.

---

51 Ebenda, S. 250 ff.
52 D. L. STUFFLEBEAM, Evaluation as Enlightment for Decision Making, a. a. O.

Diese Variablen gilt es auf den drei von STUFFLEBEAM unterschiedenen interdependenten Ebenen der Evaluation (Gemeinde, Bundesstaat, Nation) zu bedenken. In diesem Modell werden vier Situationen und die entsprechenden Entscheidungsfunktionen (Planung, Programmentwicklung, Implementierung, Modifizierung) benannt, für die vier Strategien entwickelt werden[53]. Die erste zielt auf die Evaluation des *Kontextes*. Hierbei gilt es — ähnlich der ersten der fünf oben unterschiedenen Phasen der Curriculumentwicklung —, die Bedürfnisse im Vorfeld zu erheben. Die zweite Strategie, die auf die *Eingabe* (input) gerichtet ist und deren Endergebnisse die Analyse alternativer Verfahrensentwürfe ist, will Daten für Entscheidungen über die Spezifizierung von Lernzielen, Verfahren, Materialien usw. sammeln. Die dritte beabsichtigt die Erhebung von Daten, die über den *Prozeß* gewonnen werden und die zur Verbesserung des Curriculum helfen sollen. Die vierte Evaluationsstrategie zielt darauf ab, die *Gesamtergebnisse* zu messen.

Darüber hinaus hat STUFFLEBEAM versucht, die Struktur eines Evaluationsentwurfes zu entwickeln, der sechs Kategorien und jeweils einige Unterkategorien enthält. Er dürfte wegen seines formalen Charakters als Leitlinie für Evaluationsprojekte Beachtung verdienen, obwohl er nicht die unter den einzelnen Kategorien subsumierten Probleme der Wertung und der immanenten Entscheidungen reflektiert.

1. *Konzentration der Evaluation:*
   - — Identifizierung der Entscheidungsebenen, des Evaluationsschwerpunktes, der kritischen Reflektiertheit (criticality), des Zeitpunkts, der Zusammensetzung der Alternativen;
   - — Kriterien der Entscheidungssituation durch Bestimmung der Variablen des Messens und der Urteilsnormen aufstellen;
   - — Richtlinien (policies) für die Evaluation bestimmen;
2. *Sammlung von Informationen:*
   - — Bestimmung der zu sammelnden Informationsquellen,
   - — der entsprechenden Instrumente und Methoden,
   - — des zu benutzenden Verfahrens der Stichprobenauswahl,
   - — der Bedingungen und des Plans für die Informationssammlung;
3. *Organisation der Information:*
   - — Aufstellung eines Modells für die zu sammelnden Informationen,
   - — Bestimmung eines Mittels für die Kodierung, Organisation, Speicherung und Verteilung von Informationen;
4. *Analyse der Information:*
   - — Auswahl der zu verwendenden analytischen Verfahren,
   - — Bestimmung eines Mittels für die Analyse;

53 Vgl. Anm. 17.

5. *Darstellung der Information:*
   — Definition der Adressatengruppe,
   — Bestimmung von Mitteln der Informationsvermittlung,
   — Festlegung des Formats für den Evaluationsbericht,
   — Planung der Elemente für die Darstellung von Information;
6. *Verwaltung der Evaluation:*
   — Zusammenfassung des Evaluationsplanes,
   — Bestimmung der Bedürfnisse hinsichtlich Personen und Hilfsmittel,
   — Spezifizierung der Mittel, um die Richtlinien für die Durchführung der Evaluation zu treffen,
   — Evaluation der Möglichkeit des Evaluationsplanes, valide, reliable, glaubwürdige, rechtzeitige und überzeugende Information zu erhalten,
   — Spezifizierung und Festlegung von Mitteln, um den Evaluationsentwurf periodisch auf den neuesten Stand zu bringen,
   — Aufstellung eines Etats für das ganze Evaluationsprogramm[54].

IV. Das Wert- und Entscheidungsproblem in der Evaluation

Das dritte Problemfeld der Evaluationsforschung führt in das auch wissenschaftstheoretisch strittige Feld der Wertung (vs Wertfreiheit bzw. Wertneutralität). Damit kommt auch die in Deutschland geläufige Opposition von deskriptiven und präskriptiven Verfahren ins Spiel.

In der Curriculumevaluation befaßt man sich erst seit kurzem mit der Problematik der Wertimplikationen beim Urteilen und bei Entscheidungsprozessen. So kann STAKE urteilen: „Die meisten Arbeiten über die Methodologie der Evaluation erwähnen kein Verfahren, Urteilsdaten zu sammeln und zu analysieren"[55].

WESTBURY kennzeichnet das Verhältnis von Evaluation und Beschreibung: „Evaluation kann (und muß wahrscheinlich) Beschreibung einschließen, aber Beschreibung schließt nicht notwendigerweise Evaluation ein"[56]. Nach LARKIN/SHAVER liegt die Aufgabe der Evaluation neben Datensammlung und Analyse in der Berücksichtigung von Werten, Wertinhalten und -kriterien zur Beurteilung eines Curriculum.

„Ein adäquater Evaluationsentwurf fordert sowohl Verfahren, um adäquate Schätzungsdaten (assessment data) — Schätzungsentwurf — zu erhalten, als auch Verfahren, um den Wert der Schätzungsdaten abzuwägen. Diese letzteren

54 STUFFLEBEAM, Evaluation as Enlightment for Decision Making, a. a. O., S. 70.
55 R. E. STAKE, Objectives, Priorities and other Judgement Data, RER, Vol. 40, No. 2. 1970, S. 186.
56 I. WESTBURY, Curriculum Evaluation, RER, Vol. 40, No. 2, 1970, S. 241.

Verfahren sollen von nun an als *Wertanalyse* bezeichnet werden. Ein Modell für den Wertanalyseteil eines Evaluationsentwurfes schließt substantielle und Verfahrensüberlegungen ein. Die grundlegenden substantiellen Überlegungen sind der Satz von Werten, der benutzt wird, das Curriculum zu beurteilen. Die Verfahrensüberlegungen schließen ein, wie relevante Werte zu entwickeln, wie Wertkonflikte zu lösen, wie Entscheidungskriterien festzusetzen sind, um zu bestimmen, ob ein Curriculum einem Wert angemessen genügt"[57].
JENSEN entwickelt ein Verfahren, die Rationalität der Auswahl von Lernzielen unter Zugrundelegung einer bestimmten Wertposition zu belegen[58]. Auf jeden Fall sollte der Evaluierende Kenntnis von verschiedenen Schemata zur Kategorisierung von Werten haben[59], weil dadurch die Spezifizierung von Lernzielen vereinfacht wird.

1. Werte, Prioritäten, Normen. Unterschiede in den strittigen öffentlichen Fragen beruhen auch auf unterschiedlichen Auffassungen von ethischen Werten. Neben den Wertvorstellungen, die aus direkten Interessen entstehen, bilden die ethischen Werte eine wichtige Basis für die Aufstellung von Bildungszielen und Lernzielen. Auch das ist bisher in der Curriculumforschung kaum reflektiert worden, so daß relativ wenig von den Wertvorstellungen der an der Erziehung Beteiligten bekannt ist. Der verbreitete TYLERsche Weg der Lernzielbestimmung[60] berücksichtigt kaum die den Lernzielen vorgegebenen Werte und ihre kritische Sichtung. Die Hinzuziehung von Sozialwissenschaftlern für die Erforschung der Wertvorstellungen wird wiederholt gefordert[61]. In diesem Zusammenhang beginnt die Pädagogik auch die lange vernachlässigten Beziehungen zwischen Erziehungswissenschaft und Ethik wieder aufzunehmen[62]. Unter dem Gesichtspunkt der Sammlung von Informationen für die Evaluation lassen sich

---

57 A. G. LARKINS, J. P. SHAVER, a. a. O., S. 6.
58 G. E. JENSEN, The Validation of Aims for American Democratic Education, Burgess, Minneapolis 1950.
59 Z. B.: A. N. WHITEHEAD, The Aims of Education, MACMILLAN, New York 1929. PH. E. VERNON/G. W. ALLPORT, A Test for Personal Values, Journal of Abnormal and Social Psychology, Vol. 26, 1931, S. 231—48.
D. W. OLIVER/J. P. SHAVER, Teaching Public Issues in the High School, Houghton Mifflin, Boston 1966.
M. SCRIVEN, Student Values as Educational Objectives, Publication No. 124, Social Sciences Education Consortium, Univ. of Colo, Boulder 1966.
60 R. W. TYLER, Basic Principles of Curriculum and Instruction, The University of Chicago Press, Chicago and London, 29th Edition 1969 (1st Edition 1949); vgl. auch die erste große Evaluationsstudie in den USA: E. R. SMITH, R. W. TYLER, Appraising and Recording Student Progress, Adventure in American Education, Vol. 3, Harper and Brothers, New York and London 1942 (8-year study).
61 R. E. STAKE, The Countenance of Educational Evaluation, a. a. O., S. 10.
62 Im angelsächsischen Bereich z. B.: W. SELLARS and J. HOSPERS, Readings in Ethical Theory, Appleton-Century-Crofts, New York 1952. P. H. NOWELL/SMITH, Ethics, Penguin

verschiedene Arten von Wertdaten unterscheiden. Außer *Lernzielen, summativen (Wert-)Urteilen* sind noch *Prioritäten* und *Normen* als Wertdaten zu nennen. Die Bestimmung der Prioritäten hat im Vergleich zu der genauen Lernzielbestimmung wenig Beachtung gefunden[63]. Es liegt bisher kein Modell vor, um begründet festlegen zu können, was in einer Folge von Lernzielen oder Ereignissen den Vorrang hat. Hier bedarf es weiterer Theorienbildung und Forschung.

MAGUIRE hat versucht, empirisch die Beziehungen zwischen Werten, Lernzielen und Prioritäten zu untersuchen[64]. Er erhielt Werteinschätzungen von einem heterogenen Satz von Lernzielen durch Lehrer und sodann von jeweils denselben Lehrern verschiedene Urteile darüber, welchen Lernzielen sie die Priorität geben wollten. Dabei stellten sich vor allem vier verschiedene Wert-Dimensionen der Lernziele in den Augen der Lehrer heraus: Fachwissenschaftlicher Wert, motivationale Qualität, Einfachheit der Implementation und semantische Eigenschaften. Eine andere Kategorie von Beurteilungsdaten bilden „Standards", die von STAKE „als ein erwünschtes Qualitätsniveau für etwas, wie es von einer Autorität angegeben wird"[65], begriffen werden. Ihnen entspricht im deutschen Sprachgebrauch am ehesten der Begriff Norm.

Im Anschluß an STAKE lassen sich drei Situationen identifizieren, in denen der Evaluierende Daten über Werte, Prioritäten, Normen und Lernziele sammeln kann:

— die Befragung von Personen nach einem standardisierten Protokoll,
— die Beobachtung des Personenverhaltens,
— die Analyse von Curricula durch Experten[66].

Je nach Zielsetzung der Evaluation kann sich eine Kombination der Situationen empfehlen[67].

Vier empirische Methoden bieten sich an, um die Wertvorstellungen von Befragten zu ermitteln:

Books, London 1954. R. B. BRANDT, Ethical Theory, Prentice-Hall, Englewood Cliffs, N. J. 1959. J. DEWEY, Theory of Moral Life (edited by A. ISENBERG), HOLT, RINEHART and Winston, New York 1960. M. WARNOCK, Ethics Since 1900, Oxford University Press, London 1960. M. SCRIVEN, Primary Philosophy, MCGRAW-HILL, New York 1966. D. W. OLIVER/J. P. SHAVER, a. a. O. PH. FOOT, Moral Belifs, Theories of Ethics, Oxford University Press, Oxford 1967. R. S. PETERS, Ethics and Education, GEORGE ALLEN and UNWIN, London 1968[5].

63 R. E. STAKE, Objectives, Priorities and other Judgement Data, a. a. O., S. 184 ff.

64 T. O. MAGUIRE, Value Components of Teachers' Judgements of Educational Objectives, A-V Communication Review, No. 1, 1968, S. 63—86.

65 R. E. STAKE, Objectives, Priorities and other Judgements Data, a. a. O., S. 185.

66 Ebenda, S. 189.

67 Vgl. S. B. ROBINSOHN, Bildungsreform als Revision des Curriculum, Berlin 1967, S. 47—54; ders., Ein Struktur-Konzept für Curriculum-Entwicklung, Z. f. Päd., 15. Jg., 1969, Nr. 6, S. 631—653.

— Survey,
— Skalierung,
— Q-Technik,
— Semantisches Differential (Polaritätsprofil).

Survey-Befragungen empfehlen sich dort, wo gute Daten durch direkte Befragung erhalten werden können[68]. Wenn jedoch die einzelnen Fragen mehrdeutig sind oder das Ziel der Frage unklar ist, empfehlen sich stärker redundante Verfahren, wie Skalierung[69]. Eine besondere Art ihrer Verarbeitung ist die 1953 von STEPHENSON entwickelte Q-Technik, die sich zur Ermittlung von Werten eignet[70]. Das semantische Differential ist ebenfalls ein spezielles Skalierungsverfahren. Mit ihm kann festgestellt werden, welche Bewertung Menschen bestimmten Begriffen und Vorstellungen, wie z. B. „Gewaltenteilung", zuordnen. Es wurde 1957 von OSGOOD, SUCI und TANNEBAUM entwickelt[71]. Von GEIS wurde dieses Verfahren zur Evaluation eines Projektes zur Verbesserung eines Curriculum verwendet[72]. TAYLOR und MAGUIRE benutzten es, um Biologielernziele zu untersuchen[73].

68 Vgl. Anm. 40; auch: F. F. STEPHAN and P. J. MCCARTHY, Sampling Opinions, An Analysis of Survey Procedure, WILEY, New York 1958. M. TROW, Education and Survey Research in the Social Sciences (ed. by C. Y. GLOCK), Russel Sage Foundation, New York 1967; und M. TROW, Methodological Problems in the Evaluation of Innovation, Report No. 31, Center for the Study of Evaluation, University of California, Los Angeles, May 1969. O. F. FURNO, Sample Survey Designs in Education — Focus on Administrative Utilization, RER, Vol. 36, 1966, S. 552—565.

69 J. P. GUILFORD, Psychometric Methods, MCGRAW-HILL, New York 1954. S. MESSICK, The Perceived Structure of Political Relationships, Sociometry, Vol. 24, 1961, S. 270—278. C. H. COOMBS, A Theory of Data, Wiley, New York 1964. M. E. SHAW and J. M. WRIGHT, Scales for the Measurement of Attitudes, MCGRAW-HILL, New York 1967. L. GORLOW and G. A. NOLL, The Measurement of Empirically Determined Values, Educational and Psychological Measurement, Vol. 27, 1967, S. 1115—18.
D. SJOGREN, G. W. ENGLAND, R. MELTZER, The Development of an Instrument for Assessing the Personal Values of Educational Administrators, Colo State Univ., Fort Collins, 1969. D. J. DOWD and S. C. WEST, An Inventory of Measures of Affective Behavior, Improving Educational Assessment and an Inventory of Measures of Affective Behavior (ed. W. H. BEATTY), Association for Supervision and Curriculum Development (ASCD), NEA, Washington D. C. 1969, S. 90—158.

70 W. STEPHENSON, The Study of Behavior, The University of Chicago Press, Chicago 1953. J. NUNNALLY, Tests and Measurements, Assessment and Prediction, MCGRAW-HILL, New York 1959. L. W. DOWNEY, The Task of Public Education: The Perceptions of People, Midwest Administrative Center, University of Chicago, Chicago 1960. M. SONNTAG, Attitudes Toward Education and Perception of Teacher Behavior, American Educational Research Journal (AERJ), Vol. 5, 1968, S. 385—402.

71 CH. E. OSGOOD, G. J. SUCI und P. H. TANNENBAUM, The Measurement of Meaning, The University of Ill. Press, Urbana 1957. J. G. SNIDER and CH. E. OSGOOD, Semantic Differential Technique, A Sourcebook, Aldine, Chicago 1969.

72 F. JR. GEIS, The Semantic Differential Technique as a Means of Evaluating Changes in „Affect", Doctoral Diss. Harvard University, Cambridge, Mass. 1968.
Siehe auch: O. J. HARVEY et al. Teachers' Beliefs, Classroom Atmosphere and Student Behavior, American Educational Research Journal (AERJ), Vol. 5, 1968, S. 151—66.

73 P. A. TAYLOR and T. O. MAGUIRE, Perceptions of Some Objectives for a Science Curriculum, Science Education, Vol. 51, 1967, S. 488—93.

Um Personenverhalten im Bereich der Schule zu beschreiben und zu bewerten, sind in den letzten Jahren viele Beobachtungssysteme entworfen und erprobt worden. Sie haben verschiedene Ansatzpunkte und betonen unterschiedliche Aspekte. Mehrere versuchen, aus der Beobachtung des Verhaltens die ihm immanenten Werte, Normen und Ziele zu erfassen; andere begnügen sich mit der Beschreibung der Interaktionen zwischen Lehrer und Schüler oder untersuchen die Logik des Lehrens. Eine wertvolle Sammlung von Instrumenten zur Beobachtung, Beschreibung und Bewertung von Verhalten (vor allem im Unterrichtsprozeß) enthalten die Bände „Mirrors for Behavior"[74]. Sie geben eine detaillierte Übersicht über 79 Beobachtungsverfahren, so daß auf eine Darstellung einzelner Systeme hier verzichtet wird.

Für die dritte Situation empfehlen sich die Techniken der Inhaltsanalyse, mit deren Hilfe sowohl Aufschluß über den Inhalt selbst als auch über seine Beziehung zum Produzenten und zum Konsumenten geliefert werden soll[75].

2. *Der programmatische und der öffentlich-politische Bereich.* Ein Kategorisierungsschema für Wertvorstellungen, Lern- und Bildungsziele, für Prioritäten und vor allem Ergebnisse von Curricula im Hinblick auf ihren politischen Charakter liefert BERLAK[76]. Er unterscheidet zwischen *programmatischen* (programatic) und *öffentlich-politischen* (public-policy) Ergebnissen (outcomes) von Curricula. Diese Bereichsunterscheidung geschieht im Bewußtsein des eminent politischen Charakters der Erziehung. Das Politische wird erstens der Maßstab, mit dessen Hilfe die Ergebnisse des Unterrichts in solche von öffentlich-politischem Interesse und solche programmatischen Charakters unterteilt werden, zweitens der Maßstab, Wertvorstellungen von großer öffentlich-politischer Bedeutung von solchen geringerer politischer Relevanz zu unterscheiden, drittens der Maßstab für das Recht des Evaluierenden, Urteile abzugeben. Daß auch der programmatische Bereich nicht ohne politische Implikationen ist, wird bei der Unterscheidung der beiden Bereiche durch BERLAK nicht übersehen. Die programmatischen Ergebnisse von Curricula werden durch „Lernziele" angestrebt, die öffentlich-politischen durch „Bildungsziele".

74 A. SIMON and E. G. BOYER, Mirrors for Behavior, An Anthology of Observation Instruments, Research for Better Schools, Philadelphia, Vol. 1—6, 1967, Vol. 7—14 und Vol. 15, Zusammenfassung, 1970.
Vgl. auch A. A. BELLACK, Methods of Observing Classroom Behaviour of Teachers and Students, in: K. INGENKAMP, Methods for the Evaluation of Comprehensive Schools, Weinheim, Berlin, Basel 1969, S. 187—215.

75 I. POOL, Trends in Content Analysis, Univ. of Ill. Press, Urbana 1959. B. BERELSON, Content Analysis in Communications Research, Free Press, Glencoe, Ill. 1952. D. P. CARTWRIGHT, Analysis of Qualitative Material, Research Methods in the Behavioral Sciences (ed. by L. FESTINGER and D. KATZ), HOLT, RINEHART and WINSTON, New York 1953.
Vgl. auch Anm. 36.

76 H. BERLAK, Values, Goals, Public Policy and Educational Evaluation, RER, Vol. 40, No. 2, 1970, S. 261—278.

Der Evaluierende muß sich bewußt werden, ob er „Lernziele" im programmatischen Bereich oder „Bildungsziele" im öffentlich-politischen Bereich evaluiert. Danach richtet es sich, ob er beschreibt, Urteilskriterien empfiehlt oder urteilt. Dabei stellt sich das Problem, wie die Grenze zwischen den beiden Bereichen bestimmt wird. BERLAK nennt vier Kriterien zur Bestimmung des öffentlich-politischen Bereichs, die er in Form von Fragen formuliert:

1. Ändert das Programm direkt oder indirekt das Verhältnis der Macht zwischen Bürger und Staat?
2. Betrifft das Programm sofort oder später den Status einer Person oder die Macht, die sie in dem sozialen System ausüben kann?
3. Hat das Programm einen Effekt, der danach strebt, politische oder soziale Spannungen zu vermehren oder abzubauen?
4. Bewirkt das Programm einen Wechsel im Selbstverständnis oder Selbstgefühl des Individuums?[77]

Öffentlich-politische und programmatische Ergebnisse können *geplant* oder *ungeplant* und, wenn ungeplant, noch halbwegs *erwartet* oder völlig *unerwartet* auftreten. Je nach der Zugehörigkeit zum öffentlich-politischen oder programmatischen Bereich bedarf es unterschiedlicher Evaluationsstrategien, die von verschieden ausgebildeten Personen angewendet werden. Eine sorgfältige Unterscheidung der beiden Bereiche bietet sich aus politischen und finanziellen Gründen an. Soll die Erreichung oder Verfehlung von vorher formulierten Lernzielen evaluiert werden, handelt es sich im allgemeinen um den programmatischen Bereich. Werden dagegen die Wertvorstellungen und Meinungen der Eltern über einschneidende schulorganisatorische Veränderungen evaluiert, befindet man sich im öffentlich-politischen Bereich.

Für den Evaluierenden schlägt BERLAK folgendes Verhalten vor: Im programmatischen Evaluationsbereich ist der Evaluierende aufgerufen, Werturteile abzugeben. Im Bereich der öffentlich-politischen Fragen ist es beim jetzigen Wissenschaftsstand angemessen, das Schwergewicht auf die Beschreibung zu legen, da es für diesen besonderen Bereich noch keine Evaluationsexperten gibt[78]. Bei der Evaluation kann es in beiden Bereichen zu einander widersprechenden Ergebnissen kommen, mithin auch zu Konflikten, die ausgetragen werden müssen. So könnte z. B. die Evaluation von Curricula im programmatischen Bereich in den Gesamtschulen zeigen, daß die kognitiven Ergebnisse geringer sind als die in herkömmlichen Schulen. Dennoch kann die Evaluation im öffentlich-politischen Bereich zeigen, daß diese Schulart erfolgreich ist, z. B. bei der Verwirklichung des Bildungszieles der optimalen Sozialisation des einzelnen.

---

77 Ebenda, S. 263—264.
78 Ebenda, S. 274.

3. Aspekte der Urteils- und Entscheidungsproblematik der Evaluation. Umstritten ist in der pädagogischen Diskussion, ob Wertkonflikte rational lösbar sind, ob Werturteile bewiesen und widerlegt werden können, „ob ethische Urteile in objektiver Weise gerechtfertigt werden können"[79]. Sind Wertkonflikte nicht rational lösbar, so können diejenigen, die Entscheidungsmacht haben, Ergebnisse der Evaluation aufgrund ihrer Wertvorstellungen ablehnen. GEIGER[80] und SCRIVEN[81] treten entschieden für die Lösbarkeit ethischer Differenzen mittels rationaler Strategien ein, wie sie in den Sozialwissenschaften verwendet werden. BERLAK nimmt im Anschluß an STEVENSON[82] eine Zwischenposition ein. Er tritt für die Lösbarkeit einiger Aspekte von ethisch strittigen Fragen ein, andere hält er nicht für lösbar[83]. Schwierig dürfte die Bildung von Kriterien zur Zuordnung der Aspekte zu einer der Wahlkategorien sein. Die Frage nach der Rechtfertigung von Werturteilen in objektiver Weise berührt die komplizierte Problematik von Entscheidungen im curricularen Bereich. Nach einer Unterscheidung von CRONBACH und SUPPES zielt eine Gruppe von Evaluationsuntersuchungen auf ein besseres Verständnis des untersuchten Gegenstandes, eine andere auf die wissenschaftliche Vorbereitung von Entscheidungsprozessen[84].

Es besteht Übereinstimmung darüber, daß das letztere eine zentrale Aufgabe der Evaluation ist. Um so erstaunlicher ist es, daß erst seit kurzem dieses Problem ausdrücklich thematisiert wird und daß bisher kaum Untersuchungen über Entscheidungsabläufe in diesem Bereich vorliegen. Es werden deskriptive Studien der Entscheidungsabläufe im pädagogischen Bereich gebraucht, die dazu beitragen können, den Entscheidungsprozeß in curricularen Fragen durchsichtig zu machen[85]. Entscheidungsträger müssen befragt und beobachtet werden. Einige von ihnen sind fähig, den Prozeß, der zu den Entscheidungen führt, zu verbalisieren, bei anderen wird man nur versuchen, die Prinzipien und Kriterien der Urteile von ihrem Verhalten abzulesen. Beide Arten der Untersuchung können zur Entwicklung von Modellen und Kriterien der Entscheidungsabläufe führen. Darüber hinaus dürfte auch die Untersuchung von Entscheidungsabläufen unter experimentellen Bedingungen wertvolle Hilfe zur Entwicklung der-

79 Ebenda, S. 269. Es soll nicht unerwähnt bleiben, daß sich die Debatte in der BRD relativ wenig mit der Wertproblematik beschäftigt.

80 G. GEIGER, Values and Social Science, The Planning of Change (ed. by W. G. BENNIS et al.), Holt, Rinehart and Winston, New York 1961.

81 M. SCRIVEN, Primary Philosophy, McGRAW-HILL, New York 1966.

82 CH. I. STEVENSON, Ethics and Language, Yale University Press, New Haven 1944.

83 H. BERLAK, Public Policy and Educational Evaluation, a. a. O., S. 270. Vgl. dazu auch die Methoden der analytischen Philosophie und die Ergebnisse analytischer Studien über die Wertprobleme in der Erziehung. Ein Literaturverzeichnis dazu in: J. F. SOLTIS, Einführung in die Analyse pädagogischer Begriffe, Düsseldorf 1971.

84 L. J. CRONBACH und P. SUPPES, Research for Tomorrow's Schools: Disciplined Inquiry for Education, Macmillan, New York 1969.

85 M. BERLAK, Values, Goals, Public Policy and Educational Evaluation, a. a. O., S. 272—273. Vgl. auch W. EDWARDS and A. TVERSKY, Decision Making, Penguin Books, Harmondsworth, Middlesex 1967.

artiger Modelle und Kriterien bieten. Ihrer Kenntnis bedarf der Evaluierende, damit er die Ergebnisse der Evaluation so formulieren kann, daß sie den individuellen Bedingungen der Entscheidungssituation Rechnung tragen und die Informationen enthalten, die für die betreffenden Entscheidungen wichtig sind.

GOOLER hat ein Verfahren entwickelt, das einerseits es dem Entscheidungsträger erleichtert, diejenigen Informationen zu finden, die er für die Entscheidung eines komplexen Problems benötigt, das andererseits dem Evaluierenden eine Möglichkeit bietet, die Datensammlung so anzulegen, daß sie die für den Entscheidungsprozeß relevanten Daten enthält. Es umfaßt in fünf Schritten die Identifikation:

— der Gruppe von Hauptinteressenten,
— von Schlüsselfragen, die diese Gruppe hat,
— derjenigen Daten, die benötigt werden, um die Schlüsselfragen zu beantworten,
— von Quellen für diese Daten,
— von Methoden, diese Daten zu erhalten[86].

Je präziser die Vorstellungen des Entscheidungsträgers in einzelnen Punkten sind, desto größer ist die Hilfe für die Evaluierenden, die für die Entscheidung wichtigen Daten zu sammeln. Dabei empfiehlt sich eine Kooperation von Entscheidungsträgern und Evaluierenden zu einem möglichst frühen Zeitpunkt. Die hier entwickelte Methode bietet wenig mehr als eine Formalisierung des Datensammlungsprozesses, dessen Ergebnisse eine begründete Entscheidung ermöglichen. Weitere Untersuchungen zum Beitrag der Evaluation zu Entscheidungsprozessen werden dringend benötigt. Dabei muß u. a. die von STUFFLEBEAM und GOOLER nicht problematisierte Kompetenz und Legitimität der Entscheidungsträger und die Effektivität ihrer Entscheidungen im Hinblick auf ihre Folgen thematisiert werden. Auch gilt es Prinzipien und Regeln zu entwickeln, die den Entscheidungsprozeß leiten, bzw. Kriterien, aufgrund derer er beurteilt werden kann. Darüber hinaus gilt es zu untersuchen, inwieweit in der Evaluation gewonnene Daten überhaupt Einfluß auf Entscheidungen haben. Möglicherweise werden die im Rahmen des Lot-Projekts geplanten Untersuchungen zur weiteren Klärung der Beziehungen zwischen Evaluation und Entscheidungsprozessen beitragen[87].

---

86 D. D. GOOLER, Data Collection for Educational Decision-Making, Establishing Priorities, Working Paper, o. J., Center for Instructional Research and Curriculum Evaluation, Univ. of Ill., Urbana, Ill.

87 Vgl. K.-H. FLECHSIG u. a., „Probleme der Entscheidung über Lernziele (Begründung und Aufriß des Forschungsplans zum LOT-Projekt), Programmiertes Lernen, H. 1, 1970, S. 1—32. Ders., Ein erfahrungswissenschaftlich-entscheidungstheoretischer Ansatz einer Theorie der Curriculumentwicklung, Monographie VI, Arbeitsgruppe für Unterrichtsforschung der Universität Konstanz, Konstanz 1970, vor allem S. 41—76.

V. Probleme der Curriculumevaluation

Curriculumevaluation steht in den USA und in Deutschland vor gleichen Aufgaben und Problemen. Wie es noch an einer Curriculumtheorie mangelt[88], fehlt auch eine Evaluationstheorie. Im Augenblick ist es notwendig, Begriffe und Modelle, Methoden und Instrumente der Evaluation zu entwickeln, um so die immer neu und präzise zu formulierenden Zielsetzungen der Evaluation einzelner Curricula zu verwirklichen. Die Ausweitung der zu berücksichtigenden Gesichtspunkte in der Curriculumevaluation in den letzten Jahren hat noch nicht die anstehenden Probleme gelöst, vielleicht jedoch ihre Lösung ermöglicht. Manche der oben dargestellten Ansätze sind es wert, aufgegriffen und weiterentwickelt zu werden.

Die Entwicklung einer Theorie und Methodologie der Evaluation kann nur in engem Zusammenhang mit der tatsächlichen Evaluation von Curricula erfolgen. Das bedeutet: es muß wenigstens bei der Planung einiger Curricula das Schwergewicht auf ihre Evaluation gelegt werden mit dem Ziel, eine Theorie der Evaluation zu entwickeln, die Generalisationen im Rahmen der Evaluation von Curricula erlaubt. Dabei gilt es zugleich, die Aufmerksamkeit auf die Ausbildung von Wissenschaftlern zu richten, die sich der vielschichtigen Probleme der Evaluation bewußt sind und ihre zahlreichen Methoden beherrschen[89]. Im Rahmen einer zu entwickelnden Theorie und zu erarbeitender Methoden der Curriculumevaluation sind vor allem folgende Probleme zu lösen:

1. Wie ist die Beziehung zwischen Curriculumtheorie und Evaluationstheorie zu definieren?
2. Wie werden Kriterien für die Evaluation von Curricula gewonnen?
3. Wie wird das Verhältnis von Beschreibung und Werturteilen in der Evaluation bestimmt?
4. Inwieweit ist es notwendig, das Vorfeld, das Aktionsfeld und das Wirkungsfeld in den verschiedenen Phasen der Curriculumentwicklung zu evaluieren, und welche Methoden sind angemessen?
5. Welche Funktion hat Evaluation für die verschiedenen Ebenen der Entscheidung in den unterschiedlichen Phasen der Curriculumentwicklung?
6. Welchen Beitrag kann Evaluation zur Auswahl und Revision von Lernzielen, zur Entwicklung von Curricula, zur Entwicklung spezifischer Lernsituationen und -methoden, zur Optimierung des Lernens von spezifischen Schülergruppen und Individuen leisten?
7. Wie kann eine Technologie der Evaluation entwickelt werden, und welche Instrumente und Methoden muß sie ausarbeiten?

---

88 Vgl. Zum Stand d. Curriculumforschung und -entwicklung in den USA, RER, Vol. 39, No. 3, 1969.

89 The Center for the Study of Evaluation, University of California, Los Angeles, hat ein „Training Material Development Project" in Arbeit; es werden Materialien für verschiedene Adressaten zur Einführung in Probleme und Techniken der Evaluation entwickelt.